The Statistical Value Chain - a Benchmarking Checklist for Decision Makers to Evaluate Decision Support Seen from a Statistical Point-Of-View

Herrmann, Ivan Tengbjerg; Henningsen, Geraldine; Wood, Christian D. ; Blake, James I. R. ; Mortensen, Jørgen Birk; Spliid, Henrik

Published in:
International Journal of Decision Sciences

Publication date:
2013

Document Version
Publisher's PDF, also known as Version of record

Citation (APA):
Herrmann, I. T., Henningsen, G., Wood, C. D., Blake, J. I. R., Mortensen, J. B., & Spliid, H. (2013). The Statistical Value Chain - a Benchmarking Checklist for Decision Makers to Evaluate Decision Support Seen from a Statistical Point-Of-View. *International Journal of Decision Sciences*, *4*(2), 71-83.

IJDS: Vol. 4, No. 2, July-December 2013

THE STATISTICAL VALUE CHAIN – A BENCHMARKING CHECKLIST FOR DECISION MAKERS TO EVALUATE DECISION SUPPORT SEEN FROM A STATISTICAL POINT-OF-VIEW

Ivan T. Herrmann*, Geraldine Henningsen*, Christian D. Wood, James I. R. Blake** Jørgen B. Mortensen*** and Henrik Spliid******

When decisions are made, by decision makers (DMs) in private and public organizations the DMs are supported by analysts (ANs) who provide decision support to the DM. Therefore, the quality of decision support provided by the AN directly affects the quality of a DM's decision. At present, many quantitative methods exist for evaluating uncertainty – for example, Monte Carlo simulation – and such methods work very well when the AN is in full control of the data collection and model-building processes. In many cases, however, the AN is not in control of these processes. In this article we develop a simple method that a DM can employ in order to evaluate the process of decision support from a statistical point-of-view. We call this approach the "Statistical Value Chain" (SVC): a consecutive benchmarking checklist with eight steps that can be used to evaluate decision support seen from a statistical point-of-view.

Keywords: *Statistical Value Chain; Decision Theory; Benchmarking Checklist; Decision Makers; Evaluate; Decision Support; Statistics; Data Analysis; Uncertainty; Quality*

INTRODUCTION

When decisions are made by managers and top managers—the "decision makers" (DMs) are usually heavily supported by analysts (ANs) who provide decision support. This article addresses the need for a pragmatic checklist with which to gauge the quality of information supplied to DMs by ANs. Because the quality of the information supplied by the AN directly affects the quality of a DM's decision, a DM should evaluate the quality of decision support based on well-defined criteria. Criteria of significance include levels of uncertainty in the decision-support information, the cost of the analysis, and the time taken for the decision support. Improvements in just one of these criteria can lead to improvements in others, since there is a dependency between them. The focus of this article is on the reduction of uncertainty for DMs.

Uncertainty in decision support is often evaluated through the use of simulation tools, such as the Monte Carlo approach or similar methods. Such methods work well when data, model parameters, and models are correctly defined and extracted, and (if necessary) adjusted. However, these approaches cannot compensate for corrupted data, bad data extraction or incorrectly adjusted parameters, and unclear definitions in models.

This article therefore provides a methodology for evaluating uncertainty in the decision-support process from a statistical point-of-view. This approach has been named "the Statistical Value Chain" (SVC). The SVC can be considered the "correct" way of handling data in a cradle-to-grave perspective—that is, the process from the extraction of raw data to its use for decision support.

* Climate Change and Sustainable Development, System Analysis Division, Department of Management Engineering, Technical University of Denmark, Frederiksborgvej 399, 4000 Roskilde, Denmark, *corresponding author E-mail: ithe@dtu.dk*

** University of Southampton, SO17 1BJ, UK, E-mails: c.wood@southampton.ac.uk; j.i.r.blake@soton.ac.uk

*** Department of Economics, Copenhagen University, ØsterFarimagsgade 5, 1353 Copenhagen, Denmark, *E-mail: jorgen.birk.mortensen@econ.ku.dk*

**** Department of Informatics and Mathematical Modelling, Technical University of Denmark, Building 305, DK-2800, Kgs. Lyngby, Denmark, *E-mail: spliid,hs@imm.dtu.dk*

It is the goal of this article to define and develop the SVC, which in addition to allowing uncertainty evaluation of decision support can also be considered as a best practice for an AN in providing that decision support. By benchmarking real-life decision support processes using the SVC, we can trace the consequences of practices that from a statistical point-of-view lead to uncertainty, bias, and errors in the analysis used for decision support. The steps of the SVC are developed and explained here.

The authors of the present article are not aware of a similar method developed for decision support; however, each step in the SVC is based on methods and approaches that are well-defined in the statistical literature. As such, the value of the SVC is that aggregated knowledge from different areas of statistical literature is compiled into eight simple steps that can be employed as a benchmark checklist for evaluating the quality of decision-support information.

We demonstrate the SVC through an applied decision-support case, which took place in Denmark from 2009-2012 when three large organizations requested a sustainability assessment of the production and use of biodiesel. This case is described in much greater detail by Herrmann *et al.* (2012). In the rest of the article, this decision-support case is referred to as "the biodiesel case".

CASE STUDY AND BACKGROUND

This article builds on experiences from a large biodiesel project initiated in 2008 by three organizations (Novozymes A/S, Emmelev A/S, and the Danish National Advanced Technology Foundation). The goal of the project was to develop a new enzymatic biodiesel transesterification process that theoretically should be superior to conventional transesterification. In the present article we consider these three organizations as being the DMs, or just a single DM. The Technical University of Denmark (DTU) Division of Quantitative Sustainability Assessment (QSA) was asked to conduct a sustainability assessment of the biodiesel, and is considered the AN providing decision support.

From the beginning of the project, it was stipulated that the sustainability assessment should be based on life cycle assessment (LCA) methodology. LCA is a quantitative approach used to assess the sustainability of products, technologies, and services based on different environmental impacts (Wenzel *et al.*, 1997; Finnveden *et al.*, 2009; EC-JRC, 2010). In LCA, all inputs of energy and raw materials and outputs of emissions in a cradle-to-grave perspective of a product are quantified and summarized into one or a number of environmental impact scores—for example, greenhouse gas emissions. As suggested by several researchers (including Weidema and Wesnæs, 1996; Heijungs and Frischknecht, 2005; and Hung and Ma, 2009), the typical recommendation for assessing uncertainty in an LCA is based on a two-step procedure: (1) expert guess on a distribution/uncertainty range, and (2) applying the Monte Carlo uncertainty simulation tool or similar tools. This works well when the statistical analysis is correctly carried out; however, expert guesses, Monte Carlo, and similar uncertainty simulation tools do not reveal when mistakes have been introduced into the analytical process (Gy, 1998; Petersen *et al.*, 2005). In the biodiesel case, it has been demonstrated that there was considerable uncertainty involved prior to decision support because of errors in the software tools used for the analysis (Herrmann *et al.*, 2013a). Many other similar examples of erroneous and uncertain analyses are presented in different references in the literature, for example by Makridakis (1998), Bezdek *et al.* (2002), Nielsen *et al.* (2007), and Mathiesen *et al.* (2009).

These cases support the need for a robust methodology designed to evaluate and reduce the level of uncertainty in decision support, and this is the motivation for developing the Statistical Value Chain. Here, we define the level of uncertainty in decision support with the following symbolic expression (1):

$$f(A, B, C) = U \tag{1}$$

"U" is the uncertainty level[1] of the statement delivered by the AN. "A" represents the resources available for the AN with regard to both time and capital[2]. By increasing the resources available for the AN, we can reduce the uncertainty level (and reducing the resources has the opposite effect, of course). "B" represents the size of the space (or *scope*) that is investigated. Given that there is a fixed amount of resources available for the AN, the uncertainty level can therefore be reduced by decreasing the size of the space that is investigated.

Finally, "C" represents the capability of the AN, and by increasing the capability of the AN the uncertainty level is also reduced. The uncertainty level of a given decision support can then be higher than, equal to, or lower than what the DM will accept, as outlined in this symbolic expression (2):

$$U > \alpha \; V\alpha \geq U \tag{2}$$

"α" = the *accepted* uncertainty level set by the DM. It is well known from the literature that different DMs have different accepted risk attitudes, and hence different accepted uncertainty levels (Royal Society, 1992; Farmer *et al.*, 1997; Simonet *et al.*, 1997; Estrin *et al.*, 2008). By keeping "A" and "B" constant, we can evaluate the uncertainty of an AN's decision support. The relation given in symbolic expression 1 is also given in the statistical literature, by Cochran (1977), Gy (1998), Crawley MJ (2005), and Petersen *et al.* (2005), among others. As a theoretical example we can consider two different questions concerning biofuels, keeping "A" and "C" constant:

Q1. What is the environmental impact of producing 10 tons of bioethanol in a specific company in Brazil today based on sugar cane?

Q2. What is the environmental impact from the total Brazilian production of bioethanol today?

The two questions differ significantly in terms of the size of the space under investigation. Q1 focuses on one specific company in Brazil, while Q2 is looking at all companies in Brazil that produce bioethanol (>100). The scope of Q2 is therefore much larger than of Q1. If the DM wants these questions to be answered with the same level of certainty, then the AN requires significantly more resources for data gathering for Q2 compared to Q1. On the other hand, if the resources for data gathering are fixed then the uncertainty of the answer to Q2 will increase significantly compared to Q1. The key assumption of this article is that the three variables *A, B,* and *C* basically drive the uncertainty. "B" is elaborated in Herrmann *et al.* (2013b), where a taxonomy has been developed that segregates and ranks the size of the space ("B") into 64 classes. "A" is given by the DM (or paying party), while "C" is the quality of the process that the AN uses for decision support (SVC). In general we use the expression *ceteris paribus*[3] to keep everything else constant except the quality of "C". By benchmarking real-life decision support cases of "C" with the SVC, we can evaluate the uncertainty (or reliability) of the given decision support.

The inspiration for naming our concept the "Statistical Value Chain" comes from both the statistical literature and business literature. The concept of an *analytical chain* is used in the statistical literature by Petersen *et al.* (2005) in "Representative Sampling for Reliable Data Analysis: Theory of Sampling". In the business literature, the concept *value chain* is frequently used, developed by Michael Porter in 1985 in his bestselling book "Competitive Advantage: Creating and Sustaining Superior Performance". From these concepts the idea of the *Statistical Value Chain* arose.

THE STATISTICAL VALUE CHAIN (SVC), A BENCHMARKING CHECKLIST FOR EVALUATING DECISION SUPPORT

The SVC is derived from Gy (1998) and Petersen *et al.* (2005), and also from more classical statistical and probability theory, such as Pitman (1993), Johnson (2005), and Montgomery (2005). It is not the goal of this article to describe in detail the steps of the statistical value chain—each step is described thoroughly in the literature, and relevant references are provided. The SVC can be operated by the DM as a benchmarking checklist (from a statistical perspective) of the process that the AN uses to make decision support. If the AN did not apply the SVC then the DM might ask what method was used instead to derive data, and how the data were aggregated to a final decision-support level.

From a decision-making point of view, decision support makes sense only when there are options to choose between. With no options, decision making and decision support is pointless (Lindley, 1985). Describing the likely impacts or consequences of different options or choices can be done only by collecting retrospective data and then interpreting these data in a prospective manner, and a reduction in the quality of retrospective data collection will in general lead to a troublesome prospective assessment. In other words, in a decision-support context there can be no sound scientific method of compensating for poor retrospective data collection and analysis, and a deterioration in the quality in each step of the SVC is likely to accumulate through the statistical value chain in terms of increased uncertainty and bias. Ultimately, this can make the final decision support problematic. The SVC is illustrated in Figure 1.

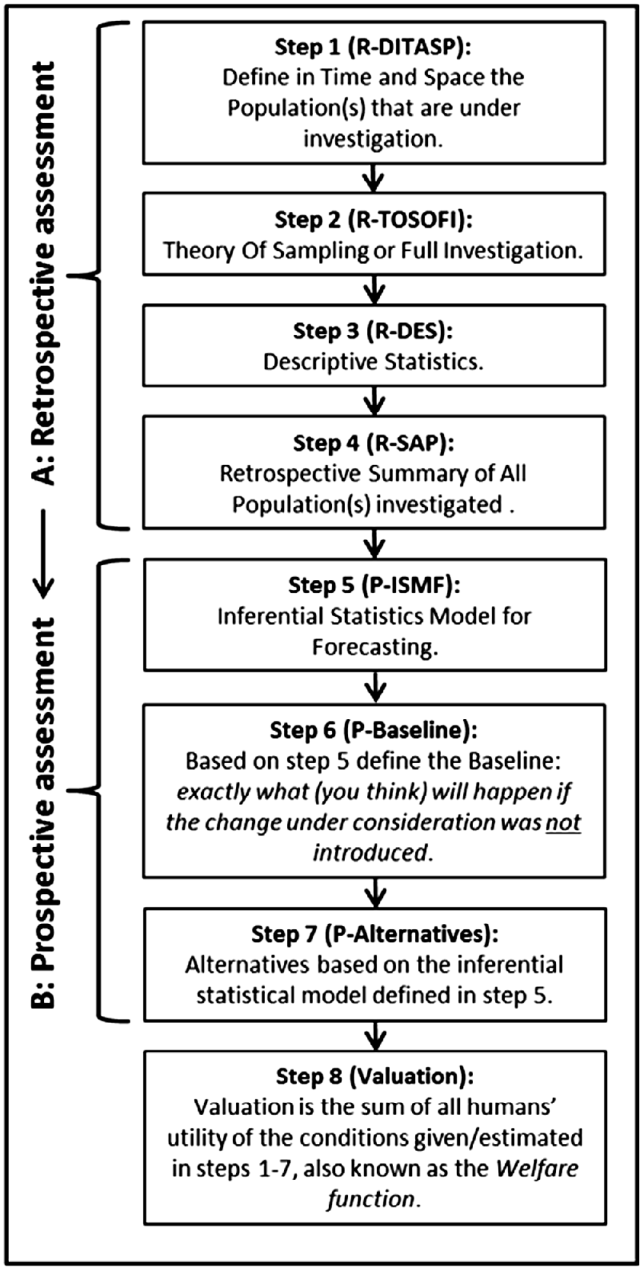

Figure 1: The Statistical Value Chain (SVC). Based on the SVC, the DM can for each step in the SVC ask whether the AN has followed suitable statistical guidelines. If not, what has the AN done instead? Each step in the SVC is essentially a check-box. The criteria for checking "yes" is simply that the DM considers the given step, based on the outline in this article, in the SVC to be sufficiently fulfilled (as given in symbolic expression 2)

Based on the SVC, the DM can for each step in the SVC ask whether the AN has followed suitable statistical guidelines. If not, what has the AN done instead? Each step in the SVC is essentially a check-box. The criteria for checking "yes" is simply that the DM considers the given step, based on the outline in this article, in the SVC to be sufficiently fulfilled (as given in symbolic expression 2).

We assume that it is a matter of *fact* what the state of the world, here denoted by S_t, is at any given point in time, where t is index in time. It is also assume that the state of the physical world can be described as the location and quantity of matter and energy in time and space.

The state of the world at $S_{t-1 \dots t-m}$ (retrospective) is unchangeable, although prospectively ($S_{t+1 \dots t+n}$) it is possible to influence the state of the world. However, it is necessary that stringent rules for induction, deduction, and abduction be applied in order to achieve the clearest picture of the state of the world and to understand how we can affect this state—that is, to change it to a more desirable state (at a later point in time). Statistics is the (applied) science of deduction, induction, and abduction, and therefore we assume that statistical analysis offers an acceptable benchmark point for evaluating the decision-support process.

Initially, we distinguish between (1) a *physical world*, which is the location and quantity of matter and energy in time and space, and (2) *value*, the worth placed on that same physical entity by one or more DMs. In the following statistical value chain, steps 1 to 7 are concerned only with the physical properties of the world.

Step 1 (R-DITASP): Defining in Time and Space the Population(s) that is/are Under Investigation

For information about the world we must collect empirical data. Obviously we cannot collect data on the entire world, but we need to collect data on the population(s) about which we are making inquiries. The starting point for this data-collecting procedure is to *define* (or outline) these populations, with regard to both space and time—for example, a specific corn field in the present year, all soybean fields in a given country in 2006, or a batch of printed circuit boards in 2028. In most decision-support contexts there are many populations from which to collect data, and we refer to this as a *system*.

Figure 2 shows a simplified model of the biodiesel case system that was investigated. The system consists of different populations, including "Rapeseed production in the field", "Rapeseed meal", and Alcohol", with a reference year of 2010.

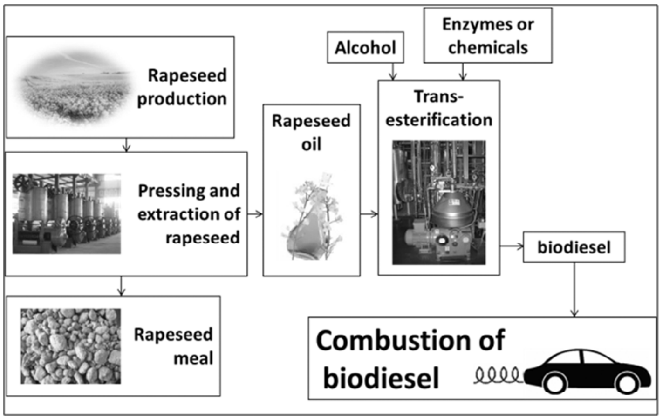

Figure 2: A simplified model of the analysed system for production and combustion of biodiesel for passenger-car transport in Denmark based on rapeseed oil. The reference year is 2010. To see the full system, refer toHerrmann *et al.* (2012)

To see the full biodiesel system, refer to Herrmann *et al.* (2012).

Based on step 1 in the SVC, the DM could ask, "Who has defined the populations to be investigated? Which populations are considered with regard to time and space? If these are not well-defined, what then?"

Step 2 (R-TOSOFI): Theory of Sampling (TOS) or Full Investigation

Once we have defined the system about which we want to make inquiries, there are two options for seeking information regarding these populations: (a) seeking full information (i.e., examining all units of the populations in the entire system), or (b) using representative sampling for each population in the system. The latter method—Theory of Sampling (TOS)—is a statistical option that if conducted properly, in principle, conserves resources compared to the first procedure where all units of the given population must be investigated fully (Cochran, 1977). Only well-used sampling procedures described by TOS can lead to representative sampling of the different populations in the product system. The starting point of any sampling procedure is outlined in step 1 of the SVC, and the sample size (and hence resources needed) depends on (1) how accurate the DM needs the results to be, (2) the size of the population, and (3) the true variation of the population. To achieve representativeness ("unbiasedness" and accuracy), it is important that all items of the sample be *randomly* chosen from the population/system, meaning that they have equal probability of being sampled[4]. For example, sampling from a batch of print circuit boards, in (e.g., four full) containers, it is not a correct sample procedure to pick the ten circuit boards closest to doors of the first container. One possible correct procedure for sampling from such a batch would be to label all the circuit boards with consecutive numbers and then to draw randomly from these numbers. Correct sampling is not a trivial task, and both Gy (1998) and Petersen *et al.* (2005) conclude that there can be grave errors in applied sampling. As noted by Petersen *et al.*, using incorrect sampling procedures can potentially corrupt the rest of the statistical value chain used for decision support: "Without representativity in this first stage in the entire analytical chain, there is no way of ever evaluating the degree of sampling bias and sampling errors embedded in the final analytical results subjected to data analysis. It has been known for more than 50 years that the combined sampling errors typically amount to 10-100, or even as much as 100-1000 times the specific analytical errors".

For the biodiesel case, the AN was never in control of the sampling process and so had no way to ensure that the principle of randomization had been applied correctly when extracting data from the different populations. The primary data source was various databases with environmental data on the different populations identified in step 1. Some of the data extracted from these databases were adequate (based on a specific accepted uncertainty level, according to symbolic expression 2); however, some of the data were rather old, and clearly not representative for the reference year (2010). Given the resources available to the AN, there were no other options than to use these data as a "best guess" and accept increased uncertainty in the final results.

Based on step 2 in the SVC, the DM could ask, "By whom and how have the data extractions been undertaken? Has full investigation been used, or has TOS been applied with a proper randomization process? If TOS has been applied, what percentage of the population has been investigated? If not full investigation or TOS, what then?"

Step 3 (R-DES): Descriptive Statistics

Descriptive statistics is about computing averages, analysis and estimating of variations, min and max, distributions, and confidence intervals etc. for each population investigated. See Johnson (2005) for further information. This step is to some degree trivial, and its quality is closely linked with the AN's capability to undertake these computations (Gy, 1998).

In the biodiesel case, the AN did *not* do such calculations, since all data in the databases were already aggregated. In the LCA software tool "SimaPro" (Pre-sustainability, 2013), an uncertainty range for the different data points is given, and five different distributions are available: "No distribution", "Range", "Triangular", "Normal distribution", and "Lognormal distribution". According to Pre-sustainability (2013), these uncertainty distribution estimates are based largely on literature studies or expert estimations. For the final decision support to the DM in the biodiesel case, the AN took the approach that the estimate of the biodiesel environmental impact was at the lowest possible uncertainty level, given the resources available for the project and the capability of the AN.

Based on step 3 in the SVC the DM could ask, "How and by whom was the descriptive statistic step undertaken?"

Step 4 (R-SAP): The Retrospective Summary of Population(s) Investigated

"Prior to determining where we are going: we must first ascertain from whence we came" (A. Lincoln). As long as a given assessment can be categorized as *retrospective*, we assume that the assessment is just a matter of accounting, and that based on the previous steps this accounting is more or less straightforward and covers the full system—that is, all populations. This process is analogous to a company's financial statement. In Gowthorpe (2003) and Andersen *et al.* (2005), the process of how to make a financial statement and its basic assumptions are described. We assume that the better (more accurate and unbiased) the accounting, the better it can serve as a *starting point* for *prospective* assessments of a system. We also assume that the better the AN is equipped to investigate the retrospective system, the better the AN can provide prospective assessments—analogous to issues treated in "Financial Statement Analysis" (Wild, 2007). In the biodiesel case, the AN calculated that the environmental impact from driving 1000 km in a diesel-engine car running on biodiesel in 2010 (retrospective) resulted in greenhouse gas emissions of 57 kg (Herrmann *et al.*, 2012).

Based on step 4 in the SVC the DM could ask, "Who performed the retrospective investigation of the population(s) investigated? What are the "retrospective" numbers that support the estimations of prospective events in the following steps of the SVC?"

Step 5 (P-ISMF): Inferential Statistics Model for Forecasting

The prospective assessment should be based on a clear model, and inferential statistics is useful for developing such a model. Both the model and the prospective assessment must be based on the information and data gathered in the past (retrospectively). Three prospective assessment methods are outlined here.

(a) Naïve forecast method. The simplest method for making a forecast is the *naïve forecast method* (Makridakis, 1998), which assumes that the best forecast for the future is the current value (of a given time series). However, in many cases it is unlikely that a system will remain *static* over a (longer) time period. Hence, using the naïve forecast method can lead to inaccuracy and bias compared to other methods, described below. Different *forces* can affect the system, which can be divided into *exogenous* forces and *endogenous* forces. Exogenous forces are those that the DM cannot (or at least, not easily) influence—they are imposed from "the outside". Endogenous forces are controlled by the DM by making various alterations to the system. In the biodiesel case we regarded the choice of alcohol type as an endogenous force on the system, since the DM (the company owner) could choose to use different types of alcohol, such as bioethanol or petrochemical methanol. However, we regarded political forces in the biodiesel system as exogenous; political changes have affected the prices, demand, and production methods for biodiesel in Europe since 2010[5]. As such, assuming that the environmental impact of the illustrated production system (Figure 2) is the same today in 2013 as it was in 2010 is probably incorrect.

(b) Times series. It can be possible to deduce how exogenous and endogenous forces impact the system by studying *time series*, at least if time series have been recorded for both the system *and* the forces that might impact the system. Based on this information, we can attempt to make forecasts and trend analyses (Makridakis, 1998). However, the study of time series can be dangerously misleading. As an example, ten different exogenous forces might affect a system, but five of these forces might be unknown to the AN, and only two of the "known forces" might have reliable times series available. If the AN makes a correlation analysis based on a response variable (such as greenhouse gas emissions) for the system and the time series for only the two known exogenous forces, the resulting forecast can be biased and misleading—after all, it omits eight of the ten potential forces affecting the system. Also, an observed correlation between different time series does not necessarily indicate that there is a *causal relation*. We see no other way to evaluate whether an observed correlation is also an expression for an actual causality except sound human judgment.

(c) Explanatory model: Information on the different forces affecting a system might (in many cases) already be summarized and available through the literature or in the mind as a memory, although both can be biased and inaccurate, as pointed out in Kahneman (2013). Given this, we do not (necessarily) need to undertake time series studies ourselves to investigate the impact of different forces on the product system. In the following, we use "explanatory variables" interchangeably with "forces". As an example, take the endogenous force "alcohol type" in the biodiesel system shown in Figure 1. For this endogenous force, information (stoichiometry) was already available in the literature, which we used to assess how this force could impact the response variable (greenhouse gas emissions) by changing the alcohol-type input to the biodiesel system. Hence, we can (also) produce forecasts through the use of *explanatory models*, which consist of explanatory variables and response variables (Montgomery, 2005). A breakdown of the explanatory variables can be useful for achieving a better result from the forecasting process. The following breakdown is not necessarily a complete list of possibilities, but rather is a suggestion for what can be considered at least a starting point.

(i) Explanatory variables can be separated into four categories, the "(un)knowns": "The known knowns, the known unknowns, the unknown knowns, and the unknown unknowns" (Herrmann et al., 2013b). This distinction between different explanatory variables is partly also discussed by Walker et al. (2003) and Montgomery (2005), who outline an uncertainty continuum going from "statistical uncertainty" to "total ignorance".

(ii) Both endogenous and exogenous explanatory variables can affect a system, so it is important to consider both types when defining the baseline (and alternatives) while forecasting. In other words, not considering important explanatory variables potentially leads to increased uncertainty in the prospective assessment, as these variables do affect and change the system.

(iii) The PESTEL framework (Johnson et al., 2005) can be used as a further breakdown of explanatory forces or variables in the explanatory model. PESTEL is an acronym for the political, economic, sociocultural, technological, environmental, and legal variables. For further information regarding the PESTEL framework, see Johnson et al. (2005).

One of the most important factors when forecasting for decision support is that the explanatory variables must be adjusted correctly, and those that impact an investigated system must *not* be missed. If variables that induce changes in a response variable describing a system are missed, the result can be bias or too much "weight" on the applied explanatory variables. As an example, take the debate of indirect land use change (ILUC). It can be misleading when land use changes are explained as driven solely by the increased production of biofuels in other countries, as seems to be the case in articles by Searchinger *et al.* (2008) and Schmidt (2010). Other variables could also drive land use changes. Kline *et al.* (2008) list a range of other possible explanatory variables beyond a single-crop market that can potentially impact this analysis. Cultural, technological, biophysical, and economic forces could, for example, also explain changes in the land use response variable. Some of these explanatory variables can also change drastically over time (that is, the coefficient used for characterizing a given explanatory variable can change), for example the already-mentioned "political" explanatory variable. In the biodiesel case a very simple inferential statistical model was employed—for details see Herrmann *et al.* (2013c).

Based on step 5 in the SVC the DM could ask, "Who performed the prospective assessment? What kind of model was employed to perform the prospective assessment?"

Step 6 (P-Baseline): Developing the Baseline for Prospective Systems

The first step for constructing a baseline should be characterized by one question: "What will happen if the change under consideration is *not* introduced?" No forces that could potentially affect the baseline should be ignored.

In the biodiesel case, the forecast of the baseline was created by assuming that the latest retrospective data point (57 kg of greenhouse gas emissions per 1000 km of driving in an ordinary diesel-engine car) would offer the best estimate of this response variable for a similar prospective event, given the resources that were available to the AN (The *naïve forecast method*). However, it was also assumed that the forecast would be valid only for a particular short period of time, in order to avoid the likely increased uncertainty from prospective changes in the biodiesel system as time progresses.

Based on step 6 in the SVC the DM could ask, "Who developed the baseline study? Have all forces that could affect the system been included? How has the baseline for prospective events been developed?"

Step 7 (P-Alternatives): Developing Alternatives to the Baseline, Rooted in the P-ISMF Model

Any relevant alternatives to the baseline study are developed in step 7 of the SVC. Like the baseline study (step 6), important exogenous and endogenous forces influencing the alternatives should not be ignored. The forces that affected the baseline might not affect the alternatives in the same way. For example, in the biodiesel case the prospective baseline would be based on petrochemical alcohol used in the transesterification process, while an alternative to the baseline would be to use biomass-based alcohol, such as bioethanol. A tax on bioethanol would then affect the alternative differently. The difference between the baseline study and alternatives describes the potential for change.

In the biodiesel case, different alternatives were applied in addition to the baseline. These alternatives were based on a change to (1) the type of alcohol, (2) the ratio of fertilizers used (given by fertilizer/manure and assuming a fixed amount of "NKP" applied to the agriculture soil), and (3) transport distance. Based on these changes in the explanatory variables for the biodiesel system, changes in the response variables could be observed—including changes in the greenhouse gas emissions from the biodiesel system. One of these alternatives gave the highest potential improvement for the response variable of the biodiesel system that was modeled, dropping from the baseline of 57 kg of CO_2 emission per 1000 km driven in an ordinary diesel-engine car to only 31 kg per 1000 km. Whether this change is considered positive or negative in the biodiesel system depends on the value that the DM places on CO_2 emission. Putting a value on such factors is the final step (step 8, below) in the SVC.

Based on step 7 in the SVC the DM could ask, "Who developed the different alternatives? Have all forces that could affect the alternatives been included? How were the different alternatives developed?"

Step 8 (Valuation): Putting a Value on the Physical Properties Given in Steps 1-7 in the SVC

Steps 1 to 7 are concerned only with strictly physical properties of the world. In step 8, "valuation" is considered. Valuation in this article is understood as the *process* of the DM placing a value on the physical entities treated in steps 1-7 above. We assume in this article that valuation takes place in collaboration with all DMs in a society that is democratic. Further, we assume that these entities can be both tangible and intangible. In the following section we refer to such tangible and intangible entities simply as "goods". It is beyond the scope of this article to compare different methods to assign values to these goods; however, three problems (points I, II, and III below) recognized in the economic literature regarding the valuation of goods indicate why valuation from an economic perspective is *not* trivial, and why without due diligence this step can lead to (increased) uncertainty if it is included by the AN and used for decision support.

The starting point of valuation from an economic perspective is an economy where no market failures take place, and as a result resources/ goods are allocated in a *Pareto Optimal* (PO) way. Pareto optimality means that resources are allocated such that it is not possible to reallocate them in a way where someone is better off without someone else being worse off (Lindeneg, 1998).

I. When a transaction in a perfect economy (without market failures) takes place, then a price is established on a good, and this is the real price of the good. Before this transaction takes place (and potentially afterwards) the owner (or any agent in the market) might, for strategic reasons, claim that the good is worth much more to the owner (or to other agents in the market) than it was actually traded for (Lindeneg, 1998; Johnson et al., 2005). Such strategic claims are of little interest in this article. Values not adopted from actual transactions have a significant risk of being biased and uncertain.

II. Transactions of non-market goods fall victim to market failures. For example, environmental problems can be considered transactions that happen outside of a perfect market (Hanley et al., 2007). Methods to determine the valuation of non-market goods are many, but such methods will be inaccurate and biased if sufficient resources and care are not taken when applying them.

III. Values placed on different physical entities change repeatedly over time. If such changes are not reflected when the value of a particular good is given, then this can also lead to increased uncertainty and bias.

Based on these factors, we find it reasonable to assume that valuations which are not adopted from a perfect market are resource-intensive (given that a "reasonable"[6] low uncertainty level is intended), and can potentially lead to bias, and hence to incorrect decision support. Cost-benefit analysis (CBA) can be used to assess the value of a given project even when the market fails. Different methods for CBA are available, such as avoided-cost analysis, social cost-benefit analysis (SCBA), cost-effectiveness analysis (CEA), and scoring methods. Concepts such as willingness-to-pay (WTP), willingness-to-accept-compensation (WTAC), and similar measures are used in the valuation of non-market goods in a CBA. For further information, see Møller (1996), Lindeneg (1998), and Hanley *et al.* (2007).

In the biodiesel case, the AN refrained from performing a valuation of the greenhouse gas emission as one of the response variables that we used for the system. Refraining from putting a value on the physical property (greenhouse gas emission) clearly has a downside in that the DM must explicitly (or implicitly) make the valuation of the changes proposed in order to consider trade-offs between other response variables—such as respiratory inorganics emissions or the cost of developing and marketing new processes—compared to the potential reduction of greenhouse gas emission from the biodiesel system.

Based on step 8 in the SVC the DM could ask, "Who has undertaken the valuation step? Has the valuation been investigated in representative way (if the society is democratic)? In general, how has data been obtained for the valuation step? What percentage of society has been asked or investigated? If asked, how have the participators been asked?"

DEMONSTRATION OF THE SVC IN DIFFERENT DECISION-SUPPORT CONTEXTS

In the following we include three concrete cases of decision support that did not follow the statistical guidelines given above, and how that decision support resulted in significant consequences for the DM or DMs.

Exploitation of Mining Blocks, Based on Gy (1998)

This case took place in 1982 in Australia, at what was the world's second-largest copper mine. Extremely biased samples were taken by hand from cones of blast-hole cuttings, with a consequent annual loss of some $8 million. In this case, the identified error seems to be in step 2 (R-TOSOFI) in the SVC. That $8 million marks the difference between "business-as-usual" and the alternate (and, seen from a statistical point-of-view) improved way of handling data in steps 6 (P-Baseline), 7 (P-Alternatives), and 8 (Valuation) in the SVC in a retrospective perspective.

Forecasting IBM's Sales, Based on Makridakis (1998)

In 1984, IBM chairman John Opel announced that sales would double to $100 billion by 1990, while profits would continue to exhibit exponential growth. Based on this forecast, IBM hired more than 100,000 new personnal. In this article we interpret this as happening at step 7 "P-Alternatives" in the SVC, since hiring 100,000 new personnal is a change compared to the 1984 level of employees. However, things did not turn out as expected. In 1996, IBM's sales were only $72 billion, while it incurred losses of more than $13 billion in 1991, 1992, and 1993. Moreover, IBM's work force was, by the end of 1996, at about half its 1986/87 peak of 430,000. Figure 3 shows the difference between the forecast and actual developments. Figure 3 also shows the retrospective numbers that led to Opel's forcast. In this article, we assume that these numbers were statistically identified correctly. However, the mistake that Opel seemed to make was in not considering changes in the business environment—that is, the exogenous variables at both step 6 (P-Baseline) and 7 (P-Alternatives) in the SVC. In this case it could seem that Opel underestimated the exogenous forces imposed externaly by competitors to IBM; the competitors also wanted a bigger slice of the growing pie.

Yearly Evaluation of the Energy Sector in a European Country

In this case, a European Country Governmental Body (ECGB) was in charge of the yearly evaluation of a new energy-saving agreement that targeted the energy sector by increasing the number of energy-saving activities. The first problem the ECGB encountered was that neither the total population of all energy-saving activities nor the characteristics of all energy-saving activities was known to them—data on these activities were available only in aggregated form. This corresponds to a diversion from the benchmark described in step 1 (R-DTSP) in the SVC. Ignoring the problem of not having identified the population that they were sampling from, the ECGB used a two-step sampling procedure. In the first step they sampled from the population of energy producers, and in a second from producer-specific projects. However, in each year the sampling was not representative of the energy producers in the population, because the distribution in the sample deviated from the distribution of the population by the type and size of the producer—large-scale energy producers or energy producers of a certain branch were highly overrepresented in the sample. In some cases, the selection probability for units of a certain type was 25 times as high as for units of another type. Also, a large-scale energy producer had a 5 times higher

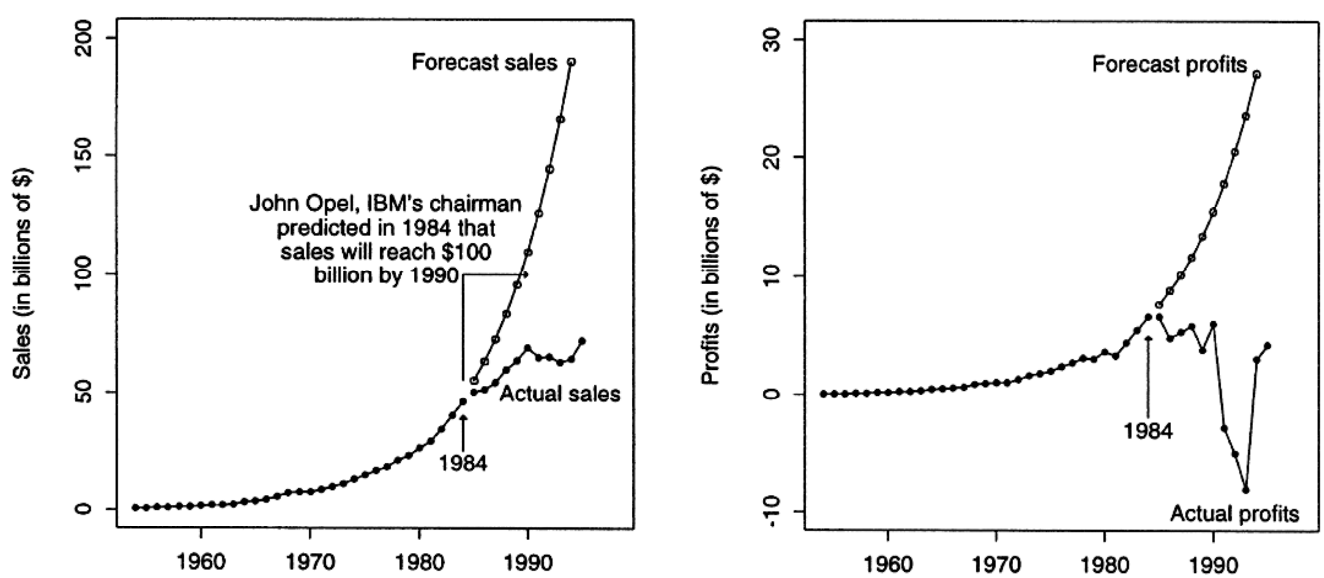

Figure 3: IBM chairman John Opel's forecast for IBM in 1984, from Makridakis (1998)

chance of being selected than a smaller energy producer, despite the fact that small-scale energy producers made up a larger share of the population (in a highly right-skewed population). Clearly the sampling procedure used by the ECGB is not in alignment with the TOS rule of equal probability for each element of the population to be sampled. The consequence of the procedure employed by the ECGB is therefore likely to be a rather biased evaluation of the energy-saving agreement, as large-scale energy producers with newer technologies were disproportionately represented in the evaluation scheme compared to smaller and older energy producers.

CONCLUDING REMARKS

In this article we have developed a benchmarking checklist, the SVC, that the DM can employ to evaluate uncertainty in decision support from a statistical perspective. Based on the SVC, the DM can for each step in the SVC ask whether the AN has followed suitable statistical guidelines. If not, what has the AN done instead? The DM can then, based on the DM's own accepted uncertainty level, decide either to reject the decision support or to continue with it. Each step in the SVC is essentially a check-box. The criteria for checking "yes" is simply that the DM considers the given step in the SVC to be sufficiently fulfilled. The value of the SVC is that it aggregates the knowledge from the statistical literature and puts it into a simple checklist that can be used by most DMs. Deviation from the statistical value chain will, at any step of the SVC, lead to increased uncertainty in the final decision support.

A challenge for the statistical value chain is that it might be relatively cost-intensive. However, both Gy (1998) and Petersen *et al.* (2005) argue that in the long run it pays off to employ proper statistical approaches when performing decision support, since it can be (and usually is) even more expensive to not use proper statistical approaches, as demonstrated with the IBM case.

The statistical value chain should not necessarily be thought of, or used as, a rigid procedure for employing statistics in decision support. As is recognized by Collins (1998), projects can rarely be put on a chain with a certain and correctly-defined number of steps before the project comes to an end. How a project develops is often better described as being an *ex-ante* "*N*-step" process, meaning that carrying out a project for decision support is an iterative activity with an unknown number of *N*-steps, going back and forth between the different steps. This is also our recommendation when using the SVC, where the fundamental principle of the Deming Circle approach ("Plan → Do → Check → Act" repeatedly) should also be used for a sound decision-support process.

ACKNOWLEDGEMENTS

We would like to thank Matthew Ritchie for proofreading a draft of this article; Dominic Roberton (Southampton University), Dan Svenstrup (Risk Analyst Danske Bank); Jens Schmidt Antonsen (Group Financial Controller ISS); Jørgen Lindgaard Petersen (Department of Management Engineering, Technical University of Denmark); Andreas Jørgensen (Department of Management Engineering, Technical University of Denmark) for their valuable and helpful comments the editors and the reviewers and editors of the International Journal of Decision Science. Funding was provided by Technical University of Denmark, Lawrence National Laboratory Berkeley, Novozymes, and The Danish National Advanced Technology Foundation.

NOTES

1. In general this would be the *expected* uncertainty level, since of course there is no guarantee that the uncertainty level will always follow this model. There can be (very few) cases for which the uncertainty level will not be as expected. To reduce the use of technical language, we use only the uncertainty level and not "the *expected* uncertainty level".
2. This would also include resources used by other parties for data gathering whothen make these data free and available to the AN.
3. In the economic literature this corresponds to the *ceteris paribus* expression used to clarify when everything else is held constant. However, in general the society of tomorrow will have accumulated more information than that of today.
4. For stratified populations, this applies to all units within a stratum.
5. In fact, the product system investigated in Herrmann *et al.* (2012) has changed significantly since 2010.
6. This is, naturally, the authors' perception of what is a reasonable uncertainty level.

REFERENCES

[1] Andersen M., Rohde C., Worre Z. (2005), Introduktiontil Virksomhedens Bogføringog Regnskab [Introduction to the Company's Accounting and

Financial Statement], 2. Edition. Samfundslitteratur, Copenhagen, Denmark.

[2] Bezdek R. H., Wendling R. M. (2002), A Half Century of Long-Range Energy Forecasts: Errors Made, Lessons Learned, and Implications for Forecasting. *Journal of Fusion Energy.* 21(3-4): 155-172.

[3] Cochran W. (1977), Sampling Techniques. 3. Edition. Wiley, New York, London.

[4] Collins D. (1998), Organizational Change: Sociological Perspectives. Routledge, New York.

[4] Crawley M. J. (2005), Statistics: an Introduction Using R. Wiley, Chichester, West Sussex, England.

[5] Estrin S., David L., Dietrich M. (2008), Microeconomics. 5. Edtion. FT Prentice Hall, Harlow.

[6] European Commission. (2010), International Reference Life Cycle Data System (ILCD) Handbook - General Guide for Life Cycle Assessment - Detailed Guidance. EUR 24708 EN. Luxembourg. Publications Office of the European Union: Joint Research Centre - Institute for Environment and Sustainability.

[7] Farmer C. M., Lund A. K., Trempel R. E., Braver E. R. (1997), Fatal Crashes of Passenger Vehicles Before and after Adding Antilock Braking Systems. Accident Analysis & Prevention. 29(6): 745-757.

[8] Finnveden G., Hauschild M. Z., Ekvall T., Guinée J., Heijungs R., Hellweg S., Koehler A., Pennington D., Suh S. (2009), Recent Developments in Life Cycle Assessment. *Journal of Environmental Management.* (91)1: 1-21.

[9] Gowthorpe C. (2003), Business Accounting and Finance: for Non-specialists. Thomson Learning, London.

[10] Gy P. (1998), Sampling for Analytical Purposes. John Wiley, Chichester, New York.

[11] Hanley N., Shogren J. F., White B. (2007), Environmental Economics: in Theory and Practice. 2. Edition. Palgrave Macmillan, Basingstoke England, New York.

[12] Hauschild M. Z. (2005), Assessing Environmental Impacts in a Life-cycle Perspective. *Environmental Science and Technology.* (39)4: 81A-88A.

[13] Heijungs, R. and R. Frischknecht. (2005), Representing Statistical Distributions for Uncertain Parameters in LCA. Relationships Between Mathematical Forms, Their Representation in EcoSpold, and Their Representation in CMLCA. *The International Journal of Life Cycle Assessment.* 10 (4): 248-54.

[14] Herrmann I. T., Jørgensen A., Bruun S., Hauschild M. Z. (2012), Potential for Optimized Production and use of Rapeseed Biodiesel. Based on a Comprehensive Real - time LCA Case Study in Denmark with Multiple Pathways. *The International Journal of Life Cycle Assessment.* (18) 2: 418-430.

[15] Herrmann I. T., Jørgensen A., Birkved M., Hauschild M. Z. (2013a), Does it Matter which LCA Tool you Choose? Comparative Assessment of SimaPro and GaBi.*Working paper*.

[16] Herrmann I. T., Hauschild M. Z., Sohn M., McKone T. (2013b), Confronting Uncertainty in Life Cycle Assessment used for Decision Support - Developing a Taxonomy for LCA studies.Journal of Industrial Ecology. *In press*.

[17] Herrmann, I. T., M. Lundberg-Jensen, A. Jørgensen, H. Spliid, T. Stidsen, and M. Z. Hauchild. (2013c), Enabling optimization in LCA: From "Ad hoc" to "Structural" LCA Approach - Based on a Biodiesel Well-to-Wheel Case Study. *The International Journal of Life Cycle Assessment.*DOI 10.1007/s11367-013-0615-z.

[18] Heijungs, R. and R. Frischknecht. (2005), Representing Statistical Distributions for Uncertain Parameters in LCA. Relationships Between Mathematical Forms, Their Representation in EcoSpold, and Their Representation in CMLCA. *The International Journal of Life Cycle Assessment.* 10 (4): 248-54.

[19] Hertwich E. G., Hammitt J. K., Pease W. S. (2000), A Theoretical Foundation for Life-Cycle Assessment. *Journal of Industrial Ecology.* 4(1): 13-28.

[20] Huijbregts MAJ, Ragas AMJ, Reijnders L, Gilijamse W. (2003), Evaluating Uncertainty in Environmental Life-Ccycle Assessment. A Case Study Comparing Two Insulation Options for a Dutch One-family Dwelling. *Environmental Science and Technology.* 37(11): 2600-2608.

[21] Hume D. (1896), A Treatise of Human Nature. Clarendon Press, Oxford.

[22] Hung, M. L., and H. Ma. (2009), Quantifying System Uncertainty of Life Cycle Assessment based on Monte Carlo Simulation. *The International Journal of Life Cycle Assessment.* 14 (1): 19-27.

[23] Johnson G., Scholes K., Whittington R. (2005), Exploring Corporate Strategy. 7. Edition. FT/Prentice Hall, Harlow, Essex, England, New York.

[24] Johnson R. A. (2005), Miller & Freund's Probability and Statistics for Engineers. 7. Edition. Pearson Prentice Hall, Upper Saddle River, NJ.

[25] Kahneman D. (2013), Thinking, Fast and Slow.1 pbkedn. New York: Farrar, Straus and Giroux.

[26] Kline K. L., Dale V. H. (2008), Biofuels: Effects on Land and Fire. Science. 321(5886).

[27] Kotter J. P. (1999), I Spidsen for Forandringer [At the Forefront of Changes]. Industriens Forlag, Peter Asschenfeldt's Nye Forlag, København.

[28] Lindeneg K. (1998), Prioritering og Styring [Prioritization and Regulation]. 2. edition. Akademisk Forlag, Copenhagen, Denmark.

[29] Lindley D. V. (1985), Making Decisions, 2. Edition. Wiley, London.

[30] Loève M. (1963), Probability Theory. 3. Edition. Van Nostrand, Princeton, N. J.

[31] Makridakis S. (1998), Forecasting: Methods and Applications. 3. edition. Wiley, New York, Chichester.

[32] Mathiesen B. V., Münster M., Fruergaard T. (2009), Uncertainties Related to the Identification of the

Marginal Energy Technology in Consequential Life Cycle Assessments. *Journal of Cleaner Production.* 17(15): 1331-1338.

[33] Montgomery D. C. (2005), Design and Analysis of Experiments. 6 edition. John Wiley & Sons, Hoboken, NJ.

[34] Møller F. (1996), Værdisætning af Miljøgoder [Valuation of Environmental Goods]. Jurist- og Økonomforbundets Forlag, København.

[35] Nielsen S. K., Karlsson K. (2007), Energy Scenarios: A Review of Methods, Uses and Suggestions for Improvement. *International Journal of Global Energy.* Issues 27(3): 302.

[36] Petersen L., Minkkinen P., Esbensen K. H. (2005), Representative Sampling for Reliable Data Analysis: Theory of Sampling. *Chemometrics and Intelligent Laboratory Systems.* 77(1-2): 261-277.

[37] Pitman J. (1993), Probability. Springer-Verlag, New York.

[38] Pre-sustainability (2013) *http://www.pre-sustainability.com/simapro-lca-software. Accessed 11* January 2013.

[39] Royal Society Great Britain (1992), Risk: Analysis, Perception and Management (Royal Society, London).

[40] Schmidt J. (2010), Comparative Life Cycle Assessment of Rapeseed Oil and Palm Oil. *The International Journal of Life Cycle Assessment.* 15(2): 183-197.

[41] Searchinger T., Heimlich R., Houghton R. A., Dong F., Elobeid A., Fabiosa J., Tokgoz S., Hayes D., Yu T. (2008), Use of US Croplands for Biofuels Increases Greenhouse Gases Through Emissions From Land-use Change. Science. 319(5867): 1238-1240.

[42] Simonet S., Wilde GJS. (1997), Risk: Perception, Acceptance and Homeostasis. *Applied Psychology - An International Review.* 46(3): 235-252.

[43] Walker W. E., Harremoës P., Rotmans J., Sluijs J. P., Asselt MBA, Janssen P., Krauss MPK. (2003), Defining Uncertainty: A Conceptual Basis for Uncertainty Management in Model-Based Decision Support. Integrated Assessment. 4(1): 5.

[44] Weick K. E. (2001), Making Sense of the Organization. Blackwell Publishers, Oxford, UK, Malden, MA.

[45] Weidema, B. and M. S. Wesnæs. (1996), Data Quality Management for Life Cycle Inventories—an Example of Using Data Quality Indicators. *Journal of Cleaner Production.* 4 (3-4): 167-74.

[46] Wenzel H., Hauschild M. Z., Alting L. (1997), Environmental Assessment of Products. Chapman & Hall, London, New York.

[47] Wild J. J. (2007), Financial Statement Analysis. 9. Edtion. McGraw-Hill/Irwin, Boston, Mass, London.